

Topological Models for Prediction of Adductability of Branched Aliphatic Compounds in Urea

SEEMA THAKRAL¹ and A. K. MADAN^{2,*}

¹GVM College of Pharmacy, Sonipat, 131-001, India; ²Faculty of Pharmaceutical Sciences, M.D. University, Rohtak, 124-001, India

(Received: 5 April 2006; in final form: 24 May 2006)

Key words: adduction, eccentric connectivity index, molecular connectivity index, topological descriptors, urea, urea inclusion compounds, Wiener's index

Abstract

The relationship of adductability of branched chain aliphatic compounds in urea with topological descriptors has been investigated. *Wiener's index* – a distance-based topological descriptor, *molecular connectivity index*, an adjacency-based topological descriptor and *eccentric connectivity index* – an adjacency-cum-distance based topological descriptor were employed for the present study. A data set comprising of 133 branched aliphatic compounds was segregated into training and test sets. The values of all the three topological indices for all the compounds constituting the training and test sets were computed using an in-house computer program. Resulting data of the training set was analyzed and suitable models were developed after identification of the adductible ranges. Subsequently, each compound in the training set was either classified as adductible or non-adductible using these models, which was then compared with the reported adductability in urea. An accuracy of prediction of $\geq 86\%$ was observed using these models in the training set. These models were then cross-validated using the test set. An accuracy of prediction of $\geq 80\%$ was observed during cross-validation of these models in an independent test set.

Introduction

Urea inclusion compounds have attracted considerable attention over the past few years because of their interesting and dynamic physicochemical properties [1–3]. Although these compounds have been known for a long time [4–6], there is still considerable interest in the inclusion phenomena varying from the complex structural behavior to basic intermolecular interactions. These inclusion compounds are conveniently considered to consist of two distinct, although not independent, host and guest substructures [7, 8]. The host substructure consists of an approximately hexagonal framework of hydrogen bonded urea with open, essentially infinite, parallel, non-intersecting tunnels (approximate diameter 5.5–5.6 Å) which completely enclose guest molecules. The guest substructure consists of guest molecules arranged in periodic repeat distance that is approximately equal to the length of guest molecule in the type of linear conformation that it must adopt in order to fit within the confined space available inside the tunnel [9, 10]. Numerous experimental as well as theoretical approaches have been used to investigate these compounds [11–19].

Though structural compatibility between host and guest components is fundamental to the most inclusion phenomena, urea forms inclusion compounds with an extremely diverse range of organic compounds [1]. Urea forms adducts under ordinary conditions with all straight chain compounds above a certain minimum length. Hence the conventional urea channel structure is formed only with guest molecules possessing a sufficiently long alkane chain and with only a limited degree of substitution of this chain allowed [20]. Linear molecules are included along the canal in an extended planar zigzag conformation. Since the urea host lattice displays a selectivity, which is essentially governed by the size of the available channel, much work has been carried out to exploit this selectivity in separating the components of mixtures on industrial scale [1]. Several empirical generalizations have been made concerning the geometric and steric features of guest molecules that form inclusion complexes with urea [21]. Thus, there is strong need to develop a mathematical model which can be relied upon to predict adductability of a branched aliphatic molecule with respect to urea as the host. Such a model may be of immense use to organic chemists for numerous applications of diverse nature and particularly to those involved in separation of isomeric compounds, utilizing urea adduction. In the present investigations, suitable models have been developed for

* Author for correspondence. E-mail: madan_ak@yahoo.com

the prediction of adductability of branched aliphatic compounds in urea.

QSPR model is a quantitative model, which relates variation in measure of activity of a series of chemical compounds to the variation in chemical structure between compounds in the series [22]. The inherent problem in development of a QSPR can be attributed to the non-quantitative nature of chemical structures. Hence, the principal objective is to find a way to represent molecules as simple numbers and to reduce the problem of structure–property relationship to a problem of correlation between two series of numbers, one of which represents the property and the other represents the chemical structure [23]. These representations are realized through the so-called molecular descriptors. Those descriptors calculated from topological representations are called topological descriptors. A topological descriptor is a single real number graph invariant characterizing a molecular graph. A graph $G = (V, E)$ is an ordered pair of two sets V and E , the former representing a non-empty set and the latter representing an unordered pair of elements of set V . When V represents atoms of a molecule and element E symbolizes covalent bonds between pairs of atoms, then G becomes a molecular graph [24]. Topological indices are generally derived from a graph by which the structural formula of a molecule can be represented. Numerical indices derived from the molecular graphs are called topological indices [25]. Thus, topological indices are usually a single number or a vector, which can be used to characterize and order molecules and predict properties. Topological descriptors that have been extensively used for structure–activity/property relationship studies include Wiener's index [26], Balaban's indices [27], Hosoya's index [28], molecular connectivity index [29], Zagreb indices [30], eccentric connectivity index [31] and E-state index [32]. In the present study, three widely used topological descriptors i.e. Wiener's index – a distance based topological descriptor, molecular connectivity index – an adjacency based topological index, and eccentric connectivity index – an adjacency cum distance based topological index have been investigated for development of models for prediction of adductability of aliphatic branched compounds in urea.

Methodology

Calculation of topological indices

Molecular connectivity index

One of the pioneering topological indices, *molecular connectivity index*, is an adjacency based topological index proposed by Randic [29]. It is denoted by χ and is defined as the sum over all the edges (ij) as per following:

$$\chi = \sum_{i=1}^n \sum_{j=1}^n (V_i V_j)^{-1/2} \quad (1)$$

where V_i and V_j are the degrees of adjacent vertices i and j and n is the number of vertices in a hydrogen suppressed molecular structure. The first-order molecular connectivity index was used very extensively in various QSPR and QSAR studies [33–37].

Wiener's index

A well-known distance-based topological index, *Wiener's index* [26], is defined as half sum of the distances between all the pairs of vertices in a hydrogen-suppressed molecular graph, that is

$$W = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n P_{ij} \quad (2)$$

where P_{ij} is the length of the path that contains the least number of edges between vertex i and vertex j in graph G and n is the maximum possible number of i and j .

Eccentric connectivity index

The eccentric connectivity index [31], an adjacency-cum-distance based topological index denoted by ξ^c is defined as the summation of the product of eccentricity and the degree of each vertex in the hydrogen suppressed molecular graph having n vertices

$$\xi^c = \sum_{i=1}^n (E_i * V_i) \quad (3)$$

where V_i is the degree of vertex i , E_i is the eccentricity of the vertex i and n is the number of the vertices in graph G . The eccentricity E_i of a vertex i in a graph G is the path length from vertex i to vertex j that is farthest from i ($E_i = \max d(ij); j \in G$); the eccentric connectivity index takes into consideration the eccentricity as well as valency of the vertices in a hydrogen-suppressed graph.

The calculation of molecular connectivity index (χ), Wiener index (W) and eccentric connectivity index (ξ^c) for three isomers of octane has been exemplified in Figure 1.

Model design

A dataset comprising of 133 molecules of diverse nature was extracted from literature [1, 4–6, 9, 10, 20, 21, 38–41] for the present investigations. The dataset comprised both adductible and non-adductible compounds and included a diverse range of branched chain organic compounds i.e. isomeric hydrocarbons, halocarbons, alcohols, acids, ketones and esters. Only those compounds were included in the data set in which any non-hydrogen atom at a position other than the terminal carbon atoms substituted the parent carbon chain and hence the molecule under consideration was truly a branched one. The dataset was equally segregated into two subsets – one to constitute a training set and the second to constitute the test set. Values of χ , W and ξ^c of all the compounds in both training and test sets were computed using an in-house computer program.

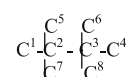
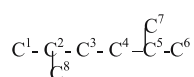
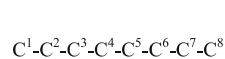
Resulting data pertaining to training set was analyzed and suitable models were developed after identification of adductible ranges by maximization of the moving average with respect to the adductible compounds ($\leq 35\%$ = non-adductible, $35\text{--}65\%$ = transitional, $\geq 65\%$ = adductible) in a manner similar to that adopted for identification of active ranges in SAR models [34]. Subsequently, each compound in the training set was classified either as adductible or non-adductible using these models, which was then compared with the reported [1, 4–6, 9, 10, 20, 21, 38–41] adductability in urea. The percentage accuracy of prediction of a particular range in the proposed model was derived from the ratio of the number of compounds classified correctly to the total number of compounds present in that range. The overall accuracy of prediction was derived from the ratio of the total number of compounds classified correctly to that of the total number of compounds present in both the adductible and non-adductible ranges.

Proposed models were cross-validated using the test set. Accordingly, each compound in the test set was classified either as adductible or non-adductible using these models, which was then compared with the reported [1, 4–6, 9, 10, 20, 21, 38–41] adductability in urea. Percent accuracy of prediction of these models was similarly determined during cross-validation.

The results are summarized in Tables 1, 2, 3 and 4.

Results and discussions

The main paradigm of chemistry is that physical, physicochemical and biological properties depend on their molecular structure. The problem in the development of a suitable correlation between chemical structures and properties can be attributed to the non-quantitative nature of chemical structures. Graph theory was successfully employed through the translation of chemical structures into characteristic numerical



Distance Matrices (D)

	1	2	3	4	5	6	7	8	E_i	S_i
1	0	1	2	3	4	5	6	7	7	28
2	1	0	1	2	3	4	5	6	6	22
3	2	1	0	1	2	3	4	5	5	18
4	3	2	1	0	1	2	3	4	4	16
5	4	3	2	1	0	1	2	3	4	16
6	5	4	3	2	1	0	1	2	5	18
7	6	5	4	3	2	1	0	1	6	22
8	7	6	5	4	3	2	1	0	7	28

	1	2	3	4	5	6	7	8	E_i	S_i
1	0	1	2	3	4	5	5	2	5	22
2	1	0	1	2	3	4	4	1	4	16
3	2	1	0	1	2	3	3	2	3	14
4	3	2	1	0	1	2	2	3	3	14
5	4	3	2	1	0	1	1	4	4	16
6	5	4	3	2	1	0	2	5	5	22
7	5	4	3	2	1	2	0	5	5	22
8	2	1	2	3	4	5	5	0	5	22

	1	2	3	4	5	6	7	8	E_i	S_i
1	0	1	2	3	2	3	2	3	3	16
2	1	0	1	2	1	2	1	2	2	10
3	2	1	0	1	2	1	2	1	2	10
4	3	2	1	0	3	2	3	2	3	16
5	2	1	2	3	0	3	2	3	3	16
6	3	2	1	2	3	0	3	2	3	16
7	2	1	2	3	2	3	0	3	3	16
8	3	2	1	2	3	2	3	0	3	16

Adjacency Matrices (A)

	1	2	3	4	5	6	7	8	V_i
1	0	1	0	0	0	0	0	0	1
2	1	0	1	0	0	0	0	0	2
3	0	1	0	1	0	0	0	0	2
4	0	0	1	0	1	0	0	0	2
5	0	0	0	1	0	1	0	0	2
6	0	0	0	0	1	0	1	0	2
7	0	0	0	0	0	1	0	1	2
8	0	0	0	0	0	0	1	1	1

	1	2	3	4	5	6	7	8	V_i
1	0	1	0	0	0	0	0	0	1
2	1	0	1	0	0	0	0	1	3
3	0	1	0	1	0	0	0	0	2
4	0	0	1	0	1	0	0	0	2
5	0	0	0	1	0	1	1	0	3
6	0	0	0	0	1	0	0	0	1
7	0	0	0	0	1	0	0	0	1
8	0	1	0	0	0	0	0	0	1

	1	2	3	4	5	6	7	8	V_i
1	0	1	0	0	0	0	0	0	1
2	1	0	1	0	1	0	1	0	4
3	0	1	0	1	0	1	0	1	4
4	0	0	1	0	0	0	0	0	1
5	0	1	0	0	0	0	0	0	1
6	0	0	1	0	0	0	0	0	1
7	0	1	0	0	0	0	0	0	1
8	0	0	1	0	0	0	0	0	1

Molecular connectivity index

$$\chi = \sum_{i=1}^n (V_i V_j)^{-1/2} = 3.914 \qquad = 3.626 \qquad = 3.25$$

Wiener index

$$W = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n P_{ij} = 84 \qquad = 74 \qquad = 58$$

Eccentric Connectivity Index

$$\xi^c = \sum_{i=1}^n (E_i * V_i) = 74 \qquad = 56 \qquad = 34$$

Figure 1. Calculation of values of molecular connectivity index (χ), Wiener index (W) and eccentric connectivity index (ξ^c) for three isomers of octane using hydrogen suppressed structure.

Table 1. Prediction of adductability of various branched aliphatic compounds in urea using proposed topological models in the training set

No.	Compound name	χ	W	ξ^c	Predicted adductability			Reported adductability
					χ	W	ξ^c	
1	3-Methyl pentane	2.83	31	29	-	-	-	-
2	2,2,4-Trimethyl pentane	3.42	66	45	-	-	-	-
3	7-Methyl tridecane	6.83	419	229	+	+	+	-
4	2,3-Dimethyl butane	2.64	29	24	-	-	-	-
5	2,2,3,4-Tetramethyl pentane	3.85	86	52	±	±	-	-
6	2,3,3-Trimethylpentane	3.52	62	28	-	-	-	-
7	2-Methyl hexane	3.27	52	47	-	-	-	-
8	2-Methyl octadecane	9.27	1124	467	+	+	+	+
9	3-Methyl tridecane	6.83	437	234	+	+	+	+
10	Hexamethylethane	3.25	58	34	-	-	-	-
11	2-Propanone	1.73	9	9	-	-	-	-
12	3-Pentanone	2.83	212	139	-	±	±	-
13	3-Hexanone	3.33	52	45	-	-	-	+
14	3-Heptanone	3.83	76	63	±	±	±	+
15	3-Octanone	4.33	112	85	±	±	±	+
16	4-Nonanone	4.83	156	111	±	±	±	+
17	2-Decanone	5.27	212	139	+	±	±	+
18	4-Decanone	5.33	222	135	+	±	±	+
19	3-Methyl butanol	2.77	32	31	-	-	-	-
20	2-Heptanone	3.77	79	65	±	±	±	+
21	Geraniol	5.16	194	122	±	±	±	-
22	2-Ethyl butanol	3.35	48	36	-	-	-	-
23	Heptanoic acid	4.27	114	87	±	±	±	+
24	Nonanoic acid	5.27	212	139	+	±	±	+
25	Undecanoic acid	6.27	354	223	+	+	+	+
26	Tridecanoic acid	7.27	556	284	+	+	+	+
27	Pentadecanoic acid	8.27	811	364	+	+	+	+
28	Octadecanoic acid	9.77	1313	521	+	+	+	+
29	Linoleic acid	9.77	1313	521	+	+	+	+
30	12-Keto stearic acid	12.16	1442	546	+	+	+	+
31	Acetic acid	1.73	9	9	-	-	-	-
32	Butanoic acid	2.77	32	31	-	-	-	+
33	2-Methyl butanoic acid	3.18	46	36	-	-	-	-
34	Sebacic acid	6.63	433	224	+	±	+	+
35	4-Pentyne-3-ol	2.83	31	29	-	-	-	-
36	Citronellyl docanoate	10.56	1569	567	±	+	+	+
37	Carbon tetrachloride	2.00	16	12	-	-	-	-
38	Hexachloroethane	3.25	58	34	-	-	-	-
39	<i>t</i> -Butyl bromide	2.00	16	12	-	-	-	-
40	Isopropyl iodide	1.73	9	9	-	-	-	-
41	2,2-Difluoro octane	4.56	146	122	±	±	±	+
42	2-Bromooctane	4.27	114	87	±	±	±	-
43	Iodoform	1.73	9	9	-	-	-	-
44	Ethyl myristicate	8.83	932	411	+	+	+	+
45	Ethyl octanoate	5.83	265	165	+	±	±	+
46	<i>n</i> -Octadecyl dodecanoate	15.83	5247	1389	+	+	+	+
47	Geranyl dodecanoate	11.56	2262	693	+	+	+	+
48	3-Methylbutyl octanoate	7.16	526	254	+	+	+	+
49	3-Methylbutyl hexanoate	6.16	324	182	+	+	+	+
50	2-Ethylbutyl dodecanoate	9.74	1193	467	+	+	+	+
51	2-Ethylhexyl dodecanoate	12.74	1552	569	+	+	+	+
52	Diethyl succinate	5.72	244	148	+	±	±	+
53	Dimethyl adipate	5.72	254	152	+	±	±	+
54	Methyl 2-methyl tetradecanoate	8.72	893	392	+	+	+	+
55	Methyl decanoate	6.33	346	221	+	+	+	+

Table 1. Continued

No.	Compound name	χ	W	ξ^c	Predicted adductability			Reported adductability
					χ	W	ξ^c	
56	Methyl tetradecanoate	8.33	792	365	+	+	+	+
57	Methyl hexadecanoate	9.33	1112	465	+	+	+	+
58	Methyl undecanoate	6.83	435	237	+	+	+	+
59	Geranyl octanoate	9.56	1164	457	+	+	+	–
60	2-Methyl butyl propionate	4.72	141	96	±	±	±	–
61	2-Ethyl butyl octanoate	7.74	587	275	+	+	+	–
62	1-Methyl heptyl propionate	6.22	316	178	+	±	±	–
63	Diisopropyl adipate	7.43	581	262	+	+	+	–
64	1-Methyl heptyl acetate	5.66	253	152	+	±	±	–
65	Dimethyl fumarate	4.72	141	96	±	±	±	–
66	<i>n</i> -Dipropyl fumarate	6.72	391	122	+	+	±	+
67	<i>n</i> -Dibutyl maleate	7.11	446	225	+	+	+	+

Note: + Adductible compounds, – non-adductible compounds, ± compounds in the transitional range, where adductability could not be specifically assigned.

Table 2. Topological models for prediction of adductability of branched aliphatic compounds in urea using training set

Model index	Nature of range in the proposed model	Index value	Number of compounds in the range		Percent accuracy
			Total	Correctly predicted	
χ	Non-adductible	< 3.56	20	18	90.0%
	Transitional	3.56–5.26	11	NA ^a	NA
	Adductible	> 5.26	36	30	83.3%
W	Non-adductible	< 66.1	19	17	89.5%
	Transitional	66.1–323.9	11	NA	NA
	Adductible	> 323.9	31	27	87.1%
ξ^c	Non-adductible	< 52.1	20	18	90.0%
	Transitional	52.1–181.9	20	NA	NA
	Adductible	> 181.9	27	24	88.8%

^aNot applicable.

descriptors by resorting to graph invariants. To obtain structure–property relationships it is necessary to find appropriate representations of these compounds. Topological descriptors are molecular descriptors derived from hydrogen suppressed topological representation of the molecule and they are expected to contain important information on a molecule that can be useful in describing physicochemical properties. In the present study three commonly used topological descriptors of diverse nature have been investigated for development of models for the prediction of adductability of compounds in urea based on the molecular structure of the guest molecule.

Retrofit analysis of data (in Tables 1–4) pertaining to model based upon χ revealed that it was possible to classify correctly a total of 48 out of 56 compounds of the training set in both the adductible and non-adductible ranges. Overall accuracy of prediction was found to be ~87% with regard to adductability in urea. The adductible range had χ values of > 5.26 and the percent accuracy of prediction in this range was found to be ~83%. The non-adductible χ range had values of < 3.56. Adductability of 90% compounds in this range

was predicted correctly. Existence of a transitional range (with χ values 3.56–5.26) between adductible and non-adductible ranges was ideal as it simply indicated a gradual change in adductability. This model was subsequently *cross-validated* using an *independent test set*. A total of 44 out of 54 compounds in the test set were classified correctly in both adductible and non-adductible ranges. Hence, the overall accuracy of prediction was found to be ~82% with regard to adductability of branched aliphatic compounds in urea during cross-validation in the independent test set.

Analysis of data (in Tables 1–4) pertaining to model based upon W revealed that, it was possible to classify correctly a total of 44 out of 50 compounds of the training set in both adductible and non-adductible ranges. The overall accuracy of prediction was found to be ~88% with regard to adductability in urea. The adductible range had W values of > 323.9 for the training set and the percent accuracy of prediction was found to be ~87%. The non-adductible W range had values of < 66.1 and adductability of ~90% of compounds was predicted correctly. Existence of a transitional range (with W values 66.1–323.9) between adductible and

Table 3. Cross-validation of the proposed topological models using test set

No.	Compound name	χ	W	ξ^c	Predicted adductability			Reported adductability
					χ	W	ξ^c	
1	3-Methyl hexane	3.33	52	45	-	-	-	-
2	3-Methyl heptane	3.83	76	63	±	±	±	-
3	2,2-Dimethyl butane	2.56	28	24	-	-	-	-
4	2,2,3-Trimethyl butane	2.94	42	29	-	-	-	-
5	2,2,3-Trimethyl pentane	3.48	63	43	-	-	-	-
6	2,3,4-Trimethylpentane	3.55	65	43	±	-	-	-
7	2-Methyl tridecane	6.27	354	223	+	+	+	+
8	2-Ethyl tridecane	6.77	444	239	+	+	+	+
9	4-Methyl tridecane	6.83	428	235	+	+	+	+
10	2-Bromodecane	5.27	212	139	+	±	±	+
11	2-Pentanone	1.73	9	9	-	-	-	+
12	2-Hexanone	3.27	52	47	-	-	-	+
13	2-Heptanone	3.77	79	65	±	±	±	+
14	4-Heptanone	3.83	75	61	±	±	±	+
15	4-Octanone	4.33	128	83	±	±	±	+
16	5-Nonanone	4.83	149	129	±	±	±	+
17	3-Decanone	5.33	226	137	+	±	±	+
18	2-10 Undecadionone	6.13	344	188	+	+	+	+
19	Butan-2-ol	2.27	18	19	-	-	-	-
20	Citronellal	5.16	194	122	±	±	±	-
21	Methyl ethyl ketone	2.27	18	19	-	-	-	+
22	2-Methyl butanol	2.83	31	29	-	-	-	-
23	Octanoic acid	4.77	158	111	±	±	±	+
24	Decanoic acid	5.77	277	169	+	±	±	+
25	Didecanoic acid	6.77	444	239	+	+	+	+
26	Tetradecanoic acid	7.77	667	321	+	+	+	+
27	Hexadecanoic acid	8.77	954	415	+	+	+	+
28	Oleic acid	9.77	1313	521	+	+	+	+
29	Linolenic acid	9.77	1313	521	+	+	+	+
30	9,12-Dihydroxy stearic acid	12.57	1556	561	+	+	+	+
31	Propionic acid	2.27	18	19	-	-	-	-
32	Hexanoic acid	3.77	89	65	±	±	±	+
33	Adipic acid	4.63	151	122	±	±	±	+
34	2-Octanol	4.27	114	87	±	±	±	+
35	Penatanoic Acid	3.27	52	47	-	-	-	+
36	6-Undecanol	5.81	261	161	+	+	±	+
37	Tetrachloroethane	2.64	29	24	-	-	-	-
38	t-Butyl Iodide	2.00	16	12	-	-	-	-
39	t-Butyl Chloride	2.00	16	12	-	-	-	-
40	Chloroform	1.73	9	9	-	-	-	-
41	2-Chlorooctane	4.27	114	87	±	±	±	-
42	Bromoform	1.73	9	9	-	-	-	-
43	Isopropyl Ether	3.13	48	38	-	-	-	-
44	Ethyl N-decanoate	6.83	428	235	+	+	+	+
45	N-Octadecyl acetate	9.77	1313	521	+	+	+	+
46	N-Octadecyl Palmitate	17.83	7485	1773	+	+	+	+
47	Citronellyl octanoate	9.56	1164	457	+	+	+	+
48	4-Methylbutyl dodecanoate	9.22	1244	436	+	+	+	+
49	2-Methylbutyl decanoate	8.22	736	338	+	+	+	+
50	2-Ethylbutyl decanoate	8.74	854	365	+	+	+	+
51	2-Ethylhexyl octanoate	8.74	818	353	+	+	+	+
52	Diethyl adipate	7.22	522	254	+	+	+	+
53	Octyl acetate	5.77	277	169	+	±	±	+
54	Methyl octanoate	5.33	226	137	+	±	±	+
55	Methyl dodecanoate	7.33	538	277	+	+	+	+

Table 3. Continued

No.	Compound name	χ	W	ξ^c	Predicted adductability			Reported adductability
					χ	W	ξ^c	
56	Methyl pentadecanoate	8.83	938	415	+	+	+	+
57	Methyl octadecanoate	12.33	1526	577	+	+	+	+
58	Ethyl acetate	2.77	32	31	-	-	-	-
59	Citronellyl butyrate	7.55	586	273	+	+	+	-
60	2-Methyl butyl octanoate	7.22	496	252	+	+	+	-
61	2-Ethyl hexyl hexanoate	7.74	563	267	+	+	+	-
62	Diethyl methyl succinate	6.11	286	159	+	±	±	-
63	Diethyl oxalate	4.72	135	92	±	±	±	-
64	Methyl 2-methyl Octanoate	5.72	246	152	+	±	±	-
65	Diethyl fumarate	5.72	244	148	+	±	±	+
66	Diethyl maleate	6.11	286	159	+	±	±	-

Note: + Adductible compounds, - non-adductible compounds, ± compounds in the transitional range, where adductability could not be specifically assigned.

Table 4. Cross-validation of proposed models for prediction of adductability in urea using independent test set

Model index	Nature of range in the cross-validated model	Index value	Number of compounds in the range		Percent accuracy
			Total	Correctly predicted	
χ	Non-adductible	< 3.56	18	14	77.8%
	Transitional	3.56–5.26	12	NA ^a	NA
	Adductible	> 5.26	36	30	83.3%
W	Non-adductible	< 66.1	20	16	80.0%
	Transitional	66.1–323.9	21	NA	NA
	Adductible	> 323.9	25	22	88.0%
ξ^c	Non-adductible	< 52.1	19	15	78.9%
	Transitional	52.1–181.9	21	NA	NA
	Adductible	> 181.9	26	23	88.5%

^aNot applicable.

non-adductible ranges was ideal as it simply indicated a gradual change in adductability. This model was also subsequently *cross-validated* using an *independent test set*. A total of 38 out of 45 compounds in the test set were classified correctly in both adductible and non-adductible ranges. Hence, the overall accuracy of prediction was found to be 84% with regard to adductability in urea during cross-validation in the independent test set.

Similarly, retrofit analysis of data (in Tables 1–4) pertaining to model based upon ξ^c for the training set revealed correct classification of a total of 42 out of 47 branched aliphatic compounds in both the adductible and non-adductible ranges. Overall accuracy of prediction was found to be ~89% with regard to adductability in urea. The adductible range had ξ^c values of > 181.9 and the percent accuracy of prediction of this range was found to be ~89%. Adductability of 90% of compounds in the non-adductible range was predicted correctly. A transitional range with ξ^c values from 52.1 to 181.9 was observed indicating a gradual change in adductability. This model was subsequently *cross-*

validated using an *independent test set*. A total of 38 out of 45 compounds in test set were classified correctly in both adductible and non-adductible ranges. Overall accuracy of prediction was found to be ~84% with regard to adductability of branched aliphatic compounds in urea during cross-validation in the independent test set.

Topological models have been successfully developed and cross-validated. These topological models offer a vast potential for their use in prediction of adductability of branched aliphatic compounds of diverse nature, which include hydrocarbons, halocarbons, ketones, alcohols and esters. However, the use of these topological models is limited to adduction of branched aliphatic compounds and that too in urea only. These models do not take care of cyclic substituents as well. Though similar strategy can be followed for development of models for other host structures but a modified strategy will, however, be necessary for development of models for either predicting any change in host structure or for determining extent of distortion in the host lattice.

References

1. K. Takemoto and N. Sonoda: In J.W. Atwood, J.E.D. Davis, and D.D. MacNicol (eds.), *Inclusion Compounds*, Academic Press, London (1984), pp. 47–67.
2. K.D.M. Harris: *Phase Trans.* **76**, 205 (2003).
3. R. Bishop and I.G. Dance: *Topics Curr. Chem.* **149**, 139 (1989).
4. O. Redlich, C.M. Gable, A.K. Dunlop, and R.W. Millar: *J. Am. Chem. Soc.* **72**, 4155 (1950).
5. A.E. Smith: *Acta Crystallogr.* **5**, 224 (1952).
6. W.J. Zimmerschied, R.A. Dinnerstein, A.W. Weitkamp, and R.F. Marschner: *Ind. Eng. Chem* **42**, 1300 (1950).
7. K.D.M. Harris and J.M. Thomas: *J. Chem. Soc. Faraday Trans.* **86**, 2985 (1990).
8. K.D.M. Harris, I. Gameson, and J.M. Thomas: *J. Chem. Soc. Faraday Trans.* **86**, 3153 (1990).
9. W. Schlenk: *Justus Liebigs Ann. Chem.* **565**, 204 (1949).
10. D. Swern: *Ind. Eng. Chem.* **47**, 216 (1955).
11. M.H. Chao, K.D.M. Harris, B.M. Kariuki, C.L. Bauer, and B.M. Foxman: *Angew. Chem. Int. Ed.* **106**, 4032 (2002).
12. R. Lefort, B. Toudic, J. Etrillard, P. Bourges, R. Currar, and T. Breczewski: *Eur. Phys. J.* **B24**, 51 (2001).
13. Waber Th., H. Boysen, and F. Frey: *Acta Cryst.* **B56**, 132 (2000).
14. M.D. Hollingsworth, U. Werner-Zwanziger, M.E. Brown, J.D. Chaney, J.C. Huffamn, K.D.M. Harris, and P.S. Sharon: *J. Am. Chem. Soc.* **121**, 9732 (1999).
15. S.C. Mayo, T.R. Welberry, M. Brown, and A. Tarr: *J. Solid State Chem.* **141**, 437 (1998).
16. R. Frost, H. Boysen, F. Frey, and H. Jagodzinski: *J. Phys. Chem. Solids* **47**, 1089 (1986).
17. I.J. Shannon, K.D.M. Harris, A.J.O. Rennie, and M.B. Webster: *J. Chem. Soc., Faraday Trans.* **89**, 2023 (1993).
18. L. Yeo and K.D.M. Harris: *J. Chem. Soc. Faraday Trans.* **94**, 1633 (1998).
19. A. Masunov and J.J. Dannenberg: *J. Phys. Chem.* **104**, 806 (2000).
20. R.W. Schiessler and D. Flitter: *J. Am. Chem. Soc.* **74**, 1720 (1950).
21. R.A. Findlay: In H.M. Schoen and J.J. Mcketta (eds.), *New Chemical Engineering Separation Techniques*, Interscience Publishers, New York (1962), pp. 257–318.
22. A.R. Katritzky and E.V. Gordeeva: *J. Chem. Inf. Comput. Sci.* **33**, 835 (1993).
23. M. Randic: *J. Chem. Inf. Comput. Sci.* **31**, 311 (1990).
24. N. Trinajstic: *Chemical graph theory*, CRC Press, Boca Raton, FL (1983).
25. S.C. Basak, A.T. Balaban, G.D. Grunwald, and B.D. Gute: *J. Chem. Inf. Comput. Sci.* **40**, 891 (2000).
26. H. Wiener: *J. Am. Chem. Soc.* **69**, 17 (1947).
27. A.T. Balaban: *Chem. Phys. Lett.* **89**, 399 (1982).
28. H. Hosoya: *J. Chem. Doc.* **12**, 181 (1972).
29. M. Randic: *J. Am. Chem. Soc.* **97**, 6609 (1975).
30. I. Gutman and M. Randic: *Chem. Phys. Lett.* **47**, 15 (1977).
31. V. Sharma, R. Goswami, and A.K. Madan: *J. Chem. Inf. Comput. Sci.* **37**, 273 (1997).
32. K. Rose, L.H. Hall, and L.B. Kier: *J. Chem. Inf. Comput. Sci.* **42**, 651 (2002).
33. L.B. Kier, L.H. Hall, W.J. Murray, and M. Randic: *J. Pharm. Sci.* **64**, 1971 (1975).
34. S. Gupta, M. Singh, and A.K. Madan: *J. Comput. Aid Mol. Desig.* **15**(7), 671–8 (2001).
35. L.B. Kier and L.H. Hall: *Molecular Connectivity in Structure–Activity Analysis*, Wiley, New York (1986).
36. M. Karelson: *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York (2000).
37. M.V. Diudea: *QSPR/QSAR Studies by Molecular Descriptors*, Nova, Huntington, New York (2001).
38. R.P. Linstead and M. Whallwy: *J. Chem. Soc.* , 2987 (1950).
39. E.V. Truter: *J. Chem. Soc.* , 2416 (1951).
40. J. Radell, J.W. Connolly, and W.R. Cosgrove: *J. Org. Chem.* **26**, 2960 (1961).
41. B. Brodman and J. Radell: *Separat. Sci.* **2**, 139 (1967).